Molecular Nutrition & Food Research
www.mnf-journal.com

Exploring the Contribution of (Poly)phenols to the Dietary Exposome Using High Resolution Mass Spectrometry Untargeted Metabolomics

Yuan-Yuan Li, Blake Rushing, Madison Schroder, Susan Sumner, and Colin D. Kay**

Scope: This study presents a workflow to construct a Dietary Exposome Library (DEL) comprised of phytochemicals and their metabolites derived from host and gut microbiome metabolism for use in peak identification/annotation of untargeted metabolomics datasets.
Methods and Results: An evidence mapping initiative established target analytes related to the consumption of phytochemical-rich foods. Analytes were confirmed by ultra-performance liquid chromatography–mass spectrometry (UPLC-MS(n)) analysis of human biospecimens from dietary intervention studies of (poly)phenol-rich diets. One hundred and sixty six verified compounds were subsequently analyzed on an untargeted metabolomics platform to acquire chromatographic and high-resolution mass spectral data for construction of a DEL. The DEL facilitated identification/annotation of 123 metabolites associate with exposure to (poly)phenol enriched diets, which included aromatic ketones, benzoic acids, ellagic acids, caffeoylquinic acids, catecholamines, coumarins, hippuric acid, hydroxytoluenes, phenylamines, stilbenes, urolithins, valerolactones, and xanthonoids, in untargeted metabolomics datasets acquire from human plasma and urine reference materials.
Conclusions: The DEL focusing on (poly)phenols and their metabolites of dietary exposure facilitated identification/annotation of ingested food components and their associated pathways in untargeted metabolomics datasets acquired from human biospecimens. The DEL continues to expand with the aim to provide evidence-based data for dietary metabolites in exposome research and inform the development of dietary intervention strategies.

1. Introduction

Untargeted metabolomics plays a critical role in exposome research, where metabolic phenotyping is used to reveal the metabolic heterogeneity among populations, the relationship to exposures, and the biological response to treatment. Untargeted metabolomics analysis of biospecimens can result in tens of thousands of signals, and with the development of data mining technologies and access to public databases, large numbers of signals/peaks related to host and microbial metabolism, and lifetime exposures (e.g., tobacco use, medications and drugs, environmental chemicals) have been identified or annotated. However, the vast majority of the acquired signals/peaks still remain unknown, creating "metabolomics dark matter."[1] Considering the high abundance of plant matter in diets (i.e., fruits, vegetables, tea, coffee, herbs, spices, botanicals etc.), it is highly likely metabolomics dark matter is composed of signals that represent dietary phytochemicals and the metabolites produced by the host and microbial metabolism. (Poly)phenols are a logical initial target to explore metabolomics dark matter as they are among the

Y.-Y. Li, B. Rushing, M. Schroder, S. Sumner
Nutrition Research Institute
UNC Chapel Hill
500 Laureate Way, Kannapolis, NC 28081, USA
E-mail: yuanyli4@unc.edu

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/mnfr.202100922

© 2022 Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

DOI: 10.1002/mnfr.202100922

Y.-Y. Li, B. Rushing, M. Schroder, S. Sumner, C. D. Kay
North Carolina Human Health Exposure Analysis Resource (NC HHEAR) Hub
NC 28081, USA
E-mail: cdkay@ncsu.edu
C. D. Kay
Food Bioprocessing and Nutrition Sciences Department
Plants for Human Health Institute
North Carolina State University
Kannapolis, NC 28081, USA

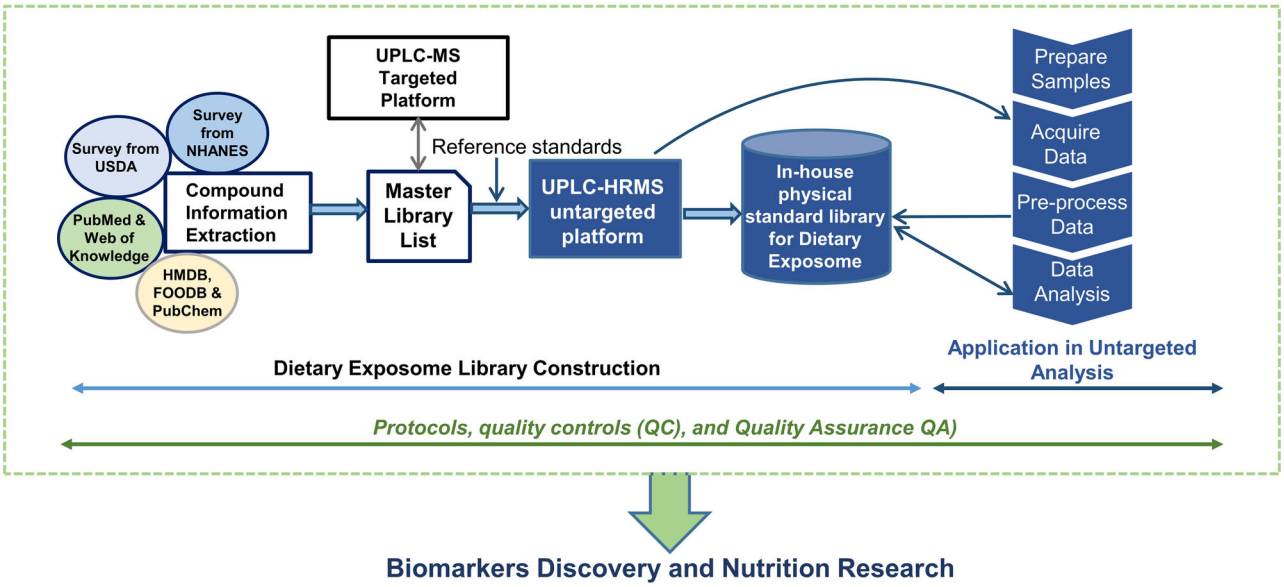

Figure 1. Workflow for Dietary Exposome Library (DEL) construction and its application in Biomarker Discovery and Nutrition Research.

most abundant phytochemicals in western diets,[2] are well characterized in foods[3] and linked to a diversity of health effects.[4] However, there is a lack of publicaly? available metabolomics databases characterizing the metabolites arising from phytochemical metabolism.

The human microbiota produces diverse small molecule phytochemical catabolites,[5] which are detected at relatively high concentrations in human plasma and urine.[6,7] Many of these molecules are structurally similar or analogous to host metabolites, including catecholamine, tyrosine, and phenylalanine pathway intermediates, which are involved in key metabolic pathways such as phase I and II metabolism or metabolism of neurotransmitters and vitamin cofactors.[8–10] Furthermore, some of these metabolites are structurally similar to drug metabolites, such as intermediates in salicylic acid and paracetamol metabolism.[11] Even though a large number of phytochemicals are included in major online databases such as PubChem and HMDB/FoodDB, many diet-related components reflect precursor structures (i.e., molecular structures present in plants/foods) which are often poorly bioavailable[12] and less likely to be detected in biospecimens (e.g., urine and plasma). Further, those bioactive and circulating phytochemical metabolites derived from host and microbial metabolism are poorly characterized, and may comprise large components of unknown signals in the untargeted metabolome. The absence of these compounds in mass spectral databases limits the ability of metabolomics researchers to discern how diet interacts with the metabolome in health and disease contexts when analyzing biospecimens. Therefore, there is a need to build and continually expand libraries that accurately reflect the circulating exposome, to aid in the identification/annotation of unknown signals in metabolomics datasets related to food ingestion. More recently, databases such as HMDB, MetaboLights, GNPS, PhytoHub, Metlin, and Biotransformer are beginning to provide greater focus on the circulating metabolites associated with the biotransformation and metabolism of dietary phytochemicals, which will be invaluable resources for expanding knowledge of the dietary exposome and aiding future nutrition and exposomics studies.

In the present manuscript we present a workflow (**Figure 1**) to construct a Dietary Exposome Library (DEL) for untargeted metabolomics, demonstrating this approach with analysis of dietary phytochemicals including their precursor form and their metabolites derived from host and gut metabolism for use in peak identification/annotation of untargeted metabolomics data. The present paper is focused on development of chemical libraries to support untargeted metabolomics efforts in Exposomics research. Others have compared the pros and cons of using high resolution MS and triple quadrupole MS in untargeted and quantitative applications.[13,14] Starting from an evidence mapping strategy utilizing systematic literature reviews, a targeted analysis by UPLC-MS(n) was used to verify the existence of selected phytochemicals in biospecimens from dietary intervention studies focused on (poly)phenol rich diets.[15–23] These verified compounds were analyzed on an untargeted metabolomic platform to acquire chromatographic and high-resolution mass spectra data for construction of a DEL. In addition, we demonstrate the ability of using the DEL to identify/annotate these (poly)phenols, their metabolites, and pathway intermediates in untargeted metabolomics datasets acquired for human plasma and urine reference materials.

2. Experimental Section

2.1. Materials

Reference standards were purchased from: Arcos Organics, (Geel, Belgium), Alfa Aesar (Tewksbury, MA, USA), Ark Pharm (Libertyville, IL, USA), Chem Impex, (Wood Dale, IL, USA), Chromadex, (Los Angeles, CA, USA), Extrasynthase SA (ZI Lyon Nord, France), Matrix Scientific (Columbia, SC, USA), Oxchem (Wood Dale, IL, USA), Polyphenols, (Sandnes, Norway), Sigma (St. Louis, MO, USA), TCI America (Portland, Oregon, USA),

© 2022 Wiley-VCH GmbH

Toronto Research Chemicals (Toronto, Canada), or were synthesized in project sponsored by the BBSRC (UK Biotechnology and Biological Sciences Research Council; BB/H004726).

2.2. Dietary Exposome Library (DEL) Construction

2.2.1. Library Target Identification

Compounds associated with the intake of (poly)phenol-rich foods were first extracted and summarized in an ongoing systematic evidence mapping initiative (search criteria Table S1, Supporting Information). The evidence map focused on bioavailable phytochemical metabolites from highly consumed (poly)phenol-rich foods, as confirmed by NHANES and USDA's FoodAPS National Household Food Acquisition and Purchase Survey,[24,25] including: coffee, tea (black, green), grain (i.e., barley, rye, wheat, rice, maze/corn), banana, apple, grape, citrus (grapefruit, orange, lime), berry (blueberry, strawberry, cranberry, raspberry), cocoa/chocolate and onion. A refined list of target compounds were verified in biospecimens collected across 11 controlled nutrition intervention studies (e.g., placebo and controlled feeding studies) that focused on consumption of foods high in (poly)phenolic compounds using LC-MS/MS[15–23] and were selected to be analyzed by an untargeted metabolomics method for construction of a DEL.

2.2.2. Library Data Acquisition on Untargeted Metabolomics Platform via Ultra-High-Performance Liquid Chromatography-High-Resolution Mass Spectrometry (UHPLC-HRMS)

Stock solutions (1 mg mL^{-1}) for each of the selected compounds (Tables S2, Supporting Information) were prepared by dissolving individual reference standards in DMSO, methanol, or water depending on solubility. Standard mixtures with a final concentration at 500 ng mL^{-1} for each compound were prepared by mixing 7-to-12 individual stock solutions with water-methanol (95:5). Isomeric compounds were prepared separately in individual solutions. A 5 µL aliquot of each standard mixture was injected for the UPLC-HRMS analysis. The DEL was composed of commercially available reference standards as indicated by the CAS numbers provided in supplement Table 2 (Table S2, Supporting Information).

Chromatographic and HRMS data were acquired on a Vanquish UHPLC system coupled to a Q Exactive HF-X Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Fisher Scientific, San Jose, CA) using conditions according to published untargeted metabolomics methods.[26,27] The chromatographic data were acquired via an HSS T3 C18 column (2.1 × 100 mm, 1.7 µm, Waters Corporation) at 50 °C with binary mobile phases of water A) and methanol B), each containing 0.1% formic acid (v/v). The linear gradient consisted of an initial composition of 2% B, increased to 100% B over 16 min, and was held at 100% B for 4 min, with a flow rate at 0.4 mL min^{-1}. The spectral data was acquired from 70 to 1050 m/z using data-dependent scanning mode. Progenesis QI (version 2.1, Waters Corporation) was used for peak picking, data extraction [retention time (RT), accurate mass (MS), and MS/MS spectral data], and construction of searchable library files. Library files also contained structural data from SDF files downloaded from Pubchem or HMDB and edited by Progenesis SDF studio.

2.3. Application of the Dietary Exposome Library in Biological Reference Materials

Urine and plasma reference materials were received from the Child Health Exposure Analysis Resource (CHEAR) consortium[28] and extracted following previously published methodology.[26] Briefly, a 50 µL aliquot of CHEAR plasma or CHEAR urine was mixed with 400-µL methanol containing 500 ng mL^{-1} L-tryptophan-d5 and vortexed at 5000 rpm for 2 min. After centrifugation at 16 000 rcf for 10 min at 4 °C, 350 µL of the supernatant was dried and reconstituted with 100 µL water-methanol (95:5, v/v). A 5 µL aliquot was injected for analysis. Untargeted metabolomics data was acquired using the instrument and method parameters described above.

The UHPLC-HRMS data from the reference materials was processed by Progenesis QI (version 2.1, Waters Corporation) for peak picking, alignment, and normalization. Peaks were matched to compounds in the Phytochemical DEL by retention time (RT), exact mass (MS), and/or MS/MS fragmentation pattern. The evidence basis for each identification was established based on an ontology level (OL) system consisting of three levels (OL1, OL2a, OL2b). Peaks that matched to compounds in the DEL by RT (±0.5 min), MS (mass error <5 ppm), and MS/MS (pattern similarity >30) were labeled as OL1, whereas peaks matching RT and MS, but not MS/MS, were labeled as OL2a. OL1 and OL2a matches were considered the highest levels of confidence. An OL2b label was provided for peaks that matched by MS and MS/MS to a compound in the Phytochemical DEL, but were outside the retention time drift tolerance (±0.5 min) compared to the purified standard.

2.4. Pathway Analysis of Detected DEL Metabolites

Compounds that were identified/annotated in the CHEAR reference urine or CHEAR reference plasma by matching with the DEL were entered into pathway analysis software, including MetaboAnalyst (www.metaboanalyst.ca)[29] and Reactome (reactome.org)[30] knowledge/databases to determine which host metabolic pathways are associated with these DEL metabolites.[31,32] KEGG identifiers for each compound were established by cross-referencing PubChem and HMDB identifiers in MetaboAnalyst (www.metaboanalyst.ca). The *p*-values and false discovery rates (FDR; Benjamini-Hochberg approach and *p*-values; binomial test) reported describe the probability of observations (metabolites) within a given pathway exceeding the number that would be randomly expected.[33]

3. Results

3.1. Library Target Identification

From our ongoing evidence mapping initiative focusing on bioavailable phytochemical metabolites from highly consumed

© 2022 Wiley-VCH GmbH

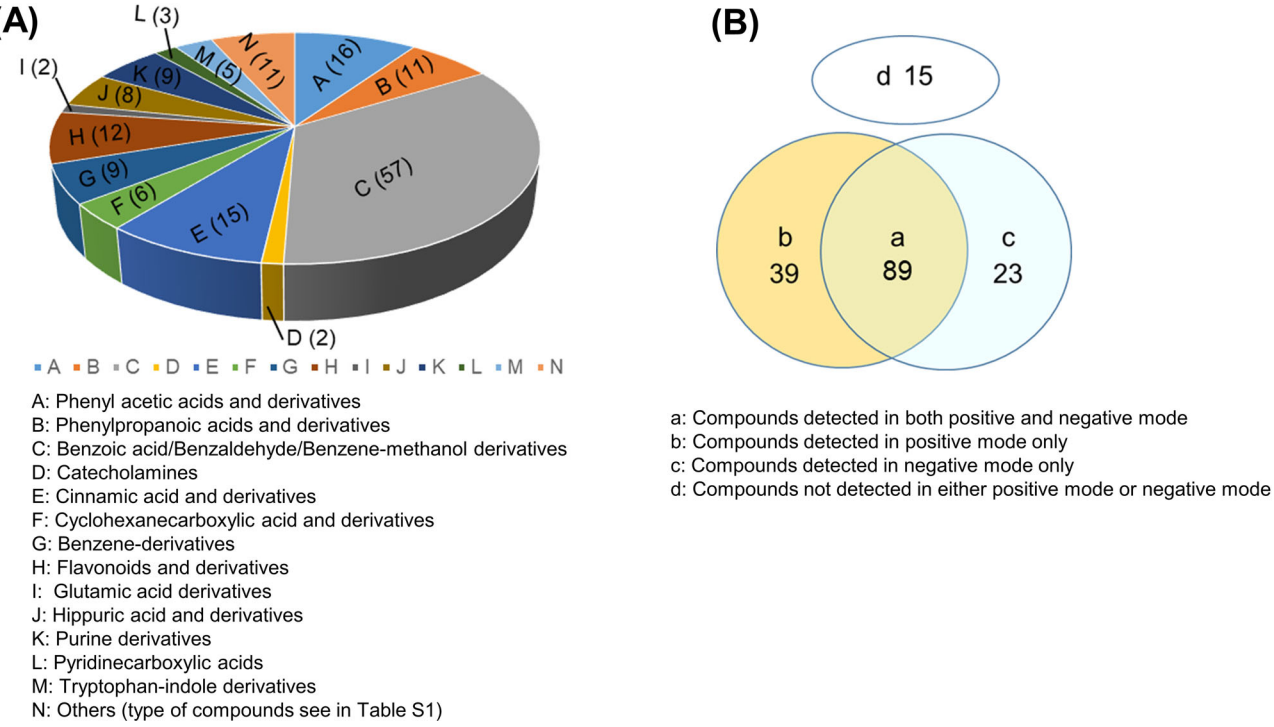

Figure 2. Selected phytochemical-metabolites for Dietary Exposome Library in the untargeted platform A) pie-chart of metabolite composition compound class; B) Venn diagram of detection results for the selected compounds.

(poly)phenol-rich foods, 166 of an initial 1029 targeted compounds were confirmed based on retention time and MS/MS spectra matching with commercially available standards, and prioritized for incorporation in the DEL. The 166 compounds were established from the total target list by their prioritization as being reported bioavailable, having available authentic reference standards, and being identified in serum or plasma of previous human nutritional intervention studies feeding polyphenol-rich foods/diets.[15–23]

3.2. Library Data Acquisition in UHPLC-HRMS Untargeted Platform

A total of 166 phytochemical/(poly)phenol related compounds (Table S2, Supporting Information) were selected to be analyzed by the UHPLC-HRMS untargeted platform (Figure 1), including 57 benzoic acids, benzaldehydes, benzene-methanol and benzene derivatives, 12 cinnamic acids, 16 phenylacetic acids, 11 (phenyl)propanoic acids, 9 purine derivatives, 8 hippuric acids, 5 tryptophan-indole derivatives, 3 pyridinecarboxylic acids, 2 catecholamines and 1 amino acid analogue, coumarin, cyclohexanecarboxylic acid, dicarboxylic acid, fatty acid conjugate, gamma amino acid, glutamic acid, phenylcarboxylic acid, hydroxy fatty acid, imidazolyl carboxylic acid, valerolactone, and a vitamin cofactor (**Figure 2**A). The majority of the metabolites (151 out of 166) were detected in either positive or negative ionization mode, with 90 compounds detected in both modes (Figure 2B). Some compounds, such as purines and tryptophan-indoles, showed strong signal responses in positive mode, which was expected due to the nitrogen-containing heterocyclic groups. In general, molecules with substituent groups (e.g., methyl, hydroxyl) in the benzene ring or the carboxylic acid side chain, and/or molecules with unsaturated carboxylic acid side chains, showed stronger signal intensities. Cyclohexanecarboxylic acid derivatives, such as chlorogenic acid (5-O-caffeoylquinic acid, 4-O-feruloylquinic acid, or 3-O-feruloylquinic acid) were not detected in the untargeted platform in either the positive or the negative mode.

3.3. Identification of Phytochemical DEL Metabolites in CHEAR Reference Urine and Plasma

After alignment and peak picking, 9708 and 11 915 signals were obtained from plasma and 22 852 and 15 230 from urine (positive and negative mode, respectively). Signals were matched with the established phytochemical DEL via exact mass, RT, and MS/MS pattern. Due to the complexity of the sample matrix, a number of molecules appeared to be structural or stereoisomers, or products of in source fragmentation, having similar accurate mass (<5 ppm tolerance) but different RT to reference standards; or could not be sufficiently resolved (**Figure 3**; an example of 3,4-dimethoxybenzoic acid). The majority of compounds (123 of 166) included in the DEL (Table S2, Supporting Information) were identified or annotated in urine and/or plasma by the untargeted platform, and were supported by evidence from three data quality ontology levels; OL1 (MS, RT, and MS/MS match), OL2a (MS and RT), or OL2b (MS and MS/MS).

© 2022 Wiley-VCH GmbH

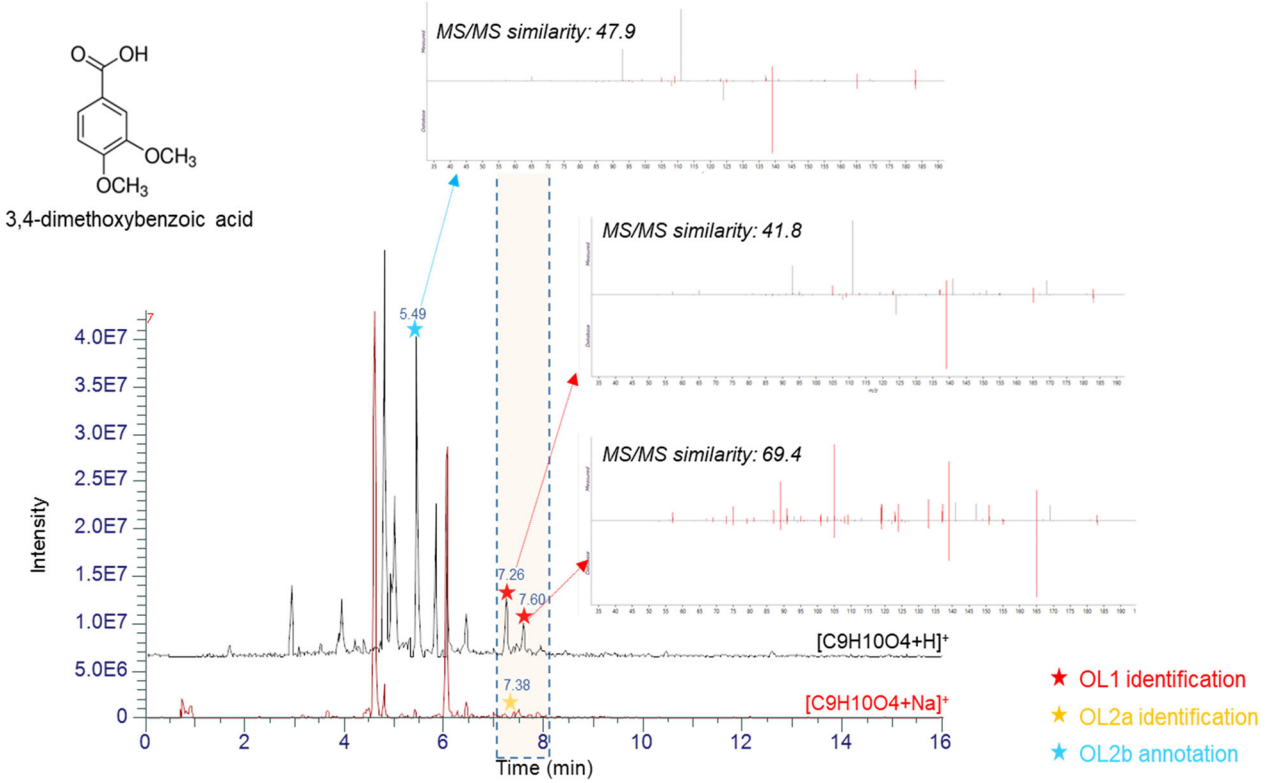

Figure 3. Identification/annotation of 3,4-dimethoxybenzoic acid (3,4-DTA) and its derivatives in urine sample. Black trace: Extracted-ion Chromatogram (EIC) of [M+H]+, Red trace: EIC of [M+Na]+. A number of peaks matched with 3,4-DTA regarding exact mass (MS, with error <5ppm). Within the retention time (RT, matching range ± 0.5 min) and tandem mass (MS/MS, profile similarity >30), two peaks (labeled in red star) satisfied the OL1 identification (MS, RT, and MS/MS). One peak (yellow star) satisfied OL2a identification, due to the matching of RT, MS but not MS/MS. One peak (blue star) satisfied OL2b annotation, due to the matching of MS, MS/MS but not RT. The peak that was annotated in OL2b level is not 3,4-DTA but a derivative related or similar to 3,4-DTA

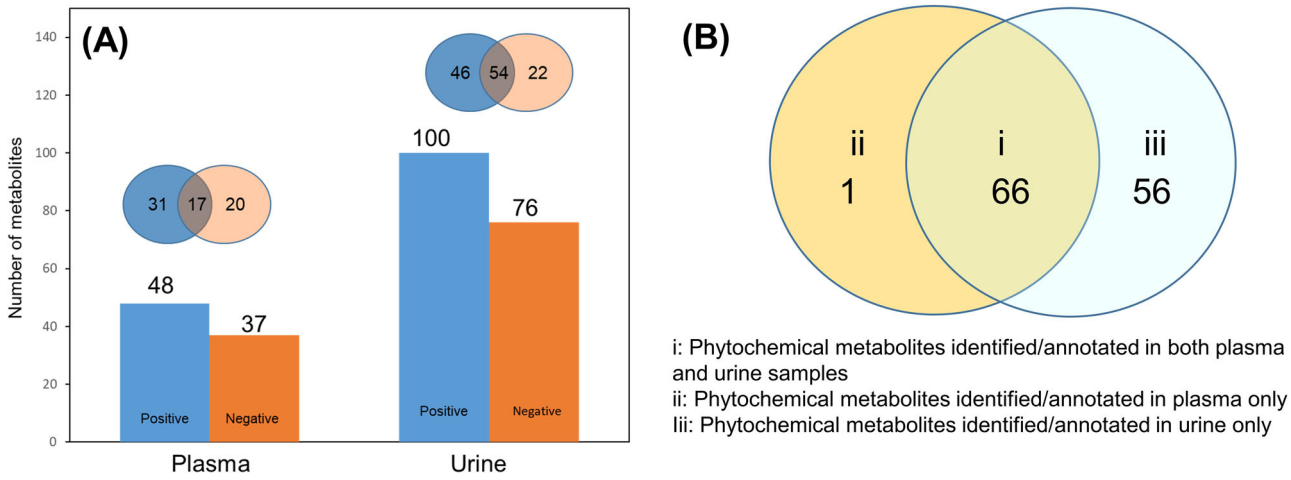

Figure 4. Phytochemical metabolites identified/annotated in CHEAR reference plasma and urine via the untargeted HRMS platform. A) Number of metabolites identified/annotated in positive and negative mode; B) number of metabolites identified/annotated in plasma and urine.

Overall, more phytochemical DEL matches were found in the CHEAR plasma/urine samples in positive mode compared to negative mode (**Figure 4**). A total of 123 metabolites from the DEL were identified/annotated across both CHEAR reference materials (Table S3, Supporting Information): 67 of the 123 metabolites were identified in both reference materials, 56 metabolites were exclusive to urine, and one metabolite was exclusive to plasma (Figure 4B). These results indicate that metabolites in the phytochemical DEL are more likely to be detected in urine samples compared to plasma. For example, aromatic alcohols and

© 2022 Wiley-VCH GmbH

aldehydes (e.g., benzaldehydes and benzene methanols) were identified/annotated from both plasma and urine samples, while aromatic acids (e.g., benzoic acids and benzeneacetic acids) were more likely to be detected in urine. Compounds that could be detected in both plasma and urine samples included purine derivatives, hippuric acid derivatives, pyridinecarboxylic acids, and tryptophan-indole derivatives. Flavonoids, including the aglycone and glycoside forms, were not found in either plasma or urine samples.

3.4. Pathway Analysis

Based on the identified/annotated phytochemical metabolites in the CHEAR reference material, we conducted pathway analysis using Reactome[30] and MetaboAnalyst[29] knowledgebase/databases, to understand the potential interaction between these phytochemicals and host metabolism. Of the 123 analytes identified in the plasma and urine CHEAR reference standards, 66 analytes had KEGG identifiers (54% coverage). Hippuric acid, methylxanthines, indoleacetates, pyridoxic acid, cataecholamines, biotin, and hydroxyphenylacetic acids mapped in Reactome, with FDR<0.1, to four major pathway categories (Table S4, Supporting Information): cell membrane, neurotransmission, metabolism and host defense. MetaboAnalyst analysis of the identified compounds having KEGG identifiers mapped to integral metabolic processes (FDR<1), including caffeine, tyrosine, phenylalanine, vitamin B6, biotin, histidine, tryptophan, arginine, proline, and purine metabolism (Table S5, Supporting Information).

4. Discussion

Metabolism is a principal driver of human health, and metabolic phenotypes (metabotypes) are influenced by a variety of internal and external factors, including genetics, diet and the environment.[34] In the present study, we investigated the ability to detect and identify metabolites associated with (poly)phenolic rich diets using an untargeted UHPLC-HR-MS platform. The phytochemicals investigated were selected from prior reports on the composition of phytochemical-dense foods and/or their metabolites post consumption, and confirmation via quantitative targeted analysis of nutrition intervention study sets (including published[15–23] and unpublished study sets). Our workflow focused on dietary metabolites that are confidently linked to ingestion of foods high in (poly)phenols, and this generalized workflow can be used to build DEL related to other food commodities. This information is critical for use in expanding public libraries and allowing researchers access to this dietary exposome data, in conjunction with big data analytics, to support present and future initiatives in nutrition for precision health.

Databases containing phytochemical metabolite data, such as HMDB, MetaboLights, GNPS, PhytoHub, Metlin, and Biotransformer are beginning to address gaps in our knowledgebase; however, significant gaps still exist for potential human/microbial biotransformation products, or chemical forms which can be empirically identified/annotated in human biospecimens following consumption of specific foods. Additionally, experimental data, particularly structurally informative tandem mass fragmentation data, are unavailable in the majority of these databases. For this reason, we have endeavored to develop and continually expand a DEL which takes these factors into consideration to aid signal identification/annotation for untargeted metabolomics data.

When discussing methodology to identify (poly)phenols, it is important to clarify that the term "(poly)phenol" refers to the collective of polyphenolic and phenolic compounds, which comprise a large subset of phytochemical subclasses, including flavonoids, stilbenes, and lignans, commonly referred to as polyphenols, and hydroxybenzoic and hydroxycinnamic acids, commonly referred to as phenolics or phenolic acids. In the literature, polyphenols such as flavonoids are most commonly reported identified using positive mode ionization, while their phenolic metabolites are most commonly identified using negative mode ionization.[35] In the present study we observed that the majority of the selected (poly)phenol metabolites were well ionized in positive mode, forming one or more adducts (e.g., $[M+H]^+$, $[M+Na]^+$, $[M+CH3OH]^+$), and demonstrated that untargeted scanning in positive mode is adequate to capture the majority of (poly)phenol metabolites (Figure 2).

One difficulty in analyzing datasets containing high numbers of small aromatic molecules is the large numbers of possible isomers and artifacts. The present analysis revealed many molecules which are either unknown structural or stereoisomers or products of in-source fragmentation, having similar accurate mass (<5 ppm tolerance) but different retention times (RT) to the library reference standards. This highlights the importance of ascribing ontology levels (i.e., data quality/certainty thresholds) when presenting HRMS evidence. Identification or annotation ontology should include exact mass, RT, and/or MS/MS spectra matches, providing evidence for the assignment of signals/peaks to metabolites and their isomers or analogues. For example, the highest level of certainty is ascribed to analytes matching accurate mass (MS), RT, and MS/MS spectra; assigned herein as ontology level 1 (OL1). Accordingly, a MS and RT match would be assigned as ontology level 2 (OL2a) (Table S3, Supporting Information). When matching the untargeted metabolomics dataset of the CHEAR reference materials against the DEL, numerous peaks had an accurate mass match (error <5ppm) for many of the phytochemical analytes, especially for those in the low mass range (150–300 Da) (Figure 3). Therefore, additional information such as RT and MS/MS spectra must be included to ensure confidence in the identification. Nevertheless, it is still difficult to achieve "one-peak-one-compound" identification, even when using RT and MS/MS information due to the complexity of biological matrices. In situations where multiple peaks match to a compound via MS, RT, and MS/MS (OL1) or MS and RT (OL2a), there is a high confidence that this compound exists within the biological sample; however, it is unclear which peak (and its corresponding intensity) is the best selection to represent that compound. In situations where one or multiple peak(s) match to a compound via MS and MS/MS but not RT, an ontology level 2b is assigned (OL2b), indicating the presence of derivative(s), isomer(s), or conjugate(s) related to the matched compound. While determining the exact chemical identity in these cases is difficult, the information obtained still provides relevant exposure and metabolic information for that compound. These cases signify the importance of reporting an evidence basis for peak identification, such as

© 2022 Wiley-VCH GmbH

the ontology system presented in this study, in addition to those required under the Metabolomics Standards Initiative (MSI).[36]

The present study describes a workflow for creating a DEL which uses experimental data from physical reference standards to aid signal identification/annotation in biospecimens using an untargeted metabolomics workflow. The focus of this study is not to investigate the ability of HRMS to improve targeted methodologies, nor to compare performances of untargeted metabolomics with targeted assays. Untargeted metabolomics is a discovery tool that simultaneously captures global signals from biospecimens without the need of a priori knowledge of its chemical composition. The generated dataset contains signal features and associated relative signal intensity that can be integrated along with sample metadata and research outcomes (e.g., states of disease, health, or wellness) to identify signals that are statistically associated with study outcomes and/or provide mechanistic insights. These signal(s) can be identified/annotated and interpreted by compound libraries, such as the DEL, and other data mining technologies. Targeted quantitative assays are normally used to then verify results from untargeted metabolomics. Even though it is recognized that the DEL has utility for targeted and MS/MS workflows, we have not provided voltages or MS/MS spectra which would support method development for such endeavors, as this is beyond the scope of the present manuscript, and more suited for a database. Further details, such as MS/MS data and chemical identifiers (InChI, SMILES etc.) will be released as part of a future online library database, which will be made publicly available.

By leveraging our established DEL, we not only provide identification for dietary metabolites based on matches with RT, MS, and MS/MS spectra (OL1 and OL2a), but also annotation for signals that shared similar MS/MS spectra with the DEL reference standards without matching with RT (OL2b). These annotations represent clusters of biologically meaningful compounds that are structurally similar to, or share significant chemical moieties with, the matched reference standard, such as conjugates, analogs, and/or derivatives, which may lead to future identifications and discoveries. In cases where investigators choose to work towards identification of unknowns (i.e., mass matches), techniques such as enzyme hydrolysis are useful in characterizing phase II conjugation, while structural elucidation can be explored using NMR, MS/MS, or chemical synthesis. A vast majority of dietary plant-based phytochemicals are poorly bioavailable in their precursor forms and are instead highly biotransformed by the gut microbiome.[5–7,12] These metabolic products are often structurally similar or analogous to host metabolites and can therefore be implicated in host metabolic processes.[8–10] In order to understand biological insight influenced by dietary phytochemicals, online pathway analysis tools such as Reactome and Metaboanalyst[29,30] were used to predict involved metabolic pathways based on the chemical identifiers (i.e., KEGG, HMDB-ID) of the identified/annotated metabolites. We have found that a number of dietary metabolites identified/annotated from CHEAR references urine and plasma by using DEL (Table S5, Supporting Information) are also included in the pathway database of Metaboanalyst and Reactome. Phenylacetic acids, purine derivatives (e.g., hypoxanthine) and tryptophan-indole derivatives (e.g., indole acetates), benzoic acids, and catecholamines, were found represented in a wide variety of biological pathways, including membrane transport, host defense, neurotransmission, and several metabolic pathways, including vitamin/xenobiotic metabolism, amino acid metabolism, purine metabolism, and phase II conjugation reactions (Tables S4,S5, Supporting Information).

Pathway databases and knowledge databases are primarily built upon genomics and proteomics data which traditionally capture endogenous enzyme and reaction data. Therefore, most pathway identifiers are associated with endogenous compounds; however, some of these compounds will also be of dietary origin. For example, hippuric acid, methylxanthines, indoleacetates, pyridoxic acid, cataecholamines, biotin, and hydroxyphenylacetic acids are mapped in Reactome to four major human pathway categories. These compounds could be considered both of endogenous and dietary origin. Hippuric acid can be derived from dietary (poly)phenols, or it could be derived endogenously by phenylalanine. Indoleacetate is also produced endogenously by tryptophan, and tryptophan is an essential amino acid which also has dietary origins. Therefore, compounds in the DEL are closely intertwined with endogenous metabolic pathways and overlap with endogenous metabolites themselves under certain circumstances. For these compounds, their effects on endogenous metabolism are more completely understood, and their implications on host metabolism can be more easily interpreted. For other compounds that are more strictly dietary (e.g., flavonoids), their interactions with host pathways are often less clear, and more studies are needed to better understand how they interact with host metabolic processes. However, research over the past several years has shown that (poly)phenols modulate the activity of endogenous metabolic pathways, including amino acid synthesis, proteostasis, oxidative pathways, autophagy, anabolic/catabolic signaling, and others.[37–40] And when phenotypic data and peak statistic information is available, these known interactions of polyphenols with endogenous metabolic pathways can be utilized to better interpret metabolic perturbations between study groups.

By leveraging our comprehensive in-house physical standard library that includes over 1000 metabolites of the host and microbial systems, in conjunction with these phytochemicals associated with Dietary Exposome, our untargeted metabolomics platform is capable of simultaneously detecting and identifying metabolites of the host and microbial systems, together with lifetime exposure including dietary exposome. Using this approach, one data capture can be used to explore the relationships between dietary intake, perturbations in host and microbial metabolism, and human health and disease. The ability to measure these compounds and detect their differences across phenotypes will significantly impact the interpretation of metabolomics data, and is a strong rationale for continuing to expand libraries to include compounds from the dietary exposome.

One of the most challenging parts of dietary exposome research lies in pathway analysis, as only a small number of phytochemical metabolites have KEGG identifiers and are therefore included in pathway mapping. In this study, 46% of the metabolites identified/annotated in the CHEAR reference urine and plasma via matching to the DEL did not have KEGG identifiers, preventing their inclusion into the pathway analysis. In some cases, these compounds may not warrant a KEGG ID, as they may not act as substrates, products, or co-factors for biochemical

© 2022 Wiley-VCH GmbH

reactions. However, in some cases, many conjugates clearly interact with metabolic pathways involving phase II detoxification reactions as they are found conjugated with sulfate and glucuronic acid, suggesting they can induce detoxification pathways.[10] In other cases, they simply may not have a KEGG ID because their contribution to biochemical reactions has not been established. Compound subclasses without KEGG identifiers were primarily benzene derivatives (benzyene diols, alcohols, aldehydes, methyl esters, catechols), hippuric acid derivatives, methoxyphenylacetic acids, and (phenyl)propanoic acids. Many of these compounds are structurally similar to known endogenous metabolites such as tryptophan/indoles, phenylalanine, and catecholamines, indicating their potential to modify metabolic pathways, likely displaying some affinity for endogenous receptors. This indicates that more efforts are needed to understand the biochemical interaction between chemicals/bio-chemicals within the dietary exposome (primarily the benzene derivatives) and host metabolism, to further increase pathway coverage in metabolomics.

With the established workflow, we will continue to expand the DEL to include more bioavailable metabolites that are associated with the dietary exposome and gradually build the DEL, including chemical IDs, molecular structures, MS and MS/MS spectra, as well as information regarding the source of exposure and bioactivity, available for public use for the analysis of untargeted data. Future research to address limitations of this study include the need for more controlled feeding trials to determine the compounds that arise in biological samples after the consumption of foods rich in specific dietary components. Additionally, mechanistic studies are required to better understand the role these compounds play in host metabolism. This knowledge will allow for a deeper understanding of the types of compounds in our diets which affect metabolic processes. Understanding the interactions between food components and host metabolism will pave the way to use food as interventions to prevent and/or manage diseases related to metabolic disorders, such as obesity, diabetes, cardiovascular diseases, and cancer, and will improve individual's quality of life throughout their lifespan.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

This study was supported by the NIEHS grant Human Health Exposure Analysis Resource (HHEAR) program through grant 1U2CES030857-01 (MPIs: Du/Fennell/Sumner) and CARDIA study 5RO1HL143885-03 (MPIs: Sumner/Gordon-Larsen/North/Avery). CDK was supported by the USDA National Institute of Food and Agriculture (Hatch/Kay-Colin; 1011757).

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Y.-Y.L. and B.R. contributed equally to this work. Conceptualization, C.D.K., S.S., and Y.-Y.L.; systematic literature review and LC-MS/MS analysis, C.D.K.; untargeted metabolomics based on UHPLC-HRMS, Y.-Y.L., B.R., and M.S.; UHPLC-HRMS library data curation and library construction, Y.-Y.L., B.R., and M.S.; biospecimen analysis using untargeted metabolomics and metabolite identification/annotation via matching against DEL, Y.-Y.L. and B.R.; pathway analysis mapping, C.D.K. and B.R.; manuscript draft, Y.-Y.L., B.R., and C.D.K.; review, editing, and production of final draft manuscript, Y.-Y.L., B.R., C.D.K., and S.S. Project supervision, S.S.

Data Availability Statement

The data that supports the findings of this study are available in the supplementary material of this article.

Keywords

dietary exposome, dietary exposome library (DEL), metabolites, polyphenols, untargeted metabolomics

Received: October 8, 2021
Revised: January 11, 2022
Published online: February 25, 2022

[1] R. R. da Silva, P. C. Dorrestein, R. A. Quinn, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 12549.
[2] M. M. Most, *J. Am. Diet. Assoc.* **2004**, *104*, 1725.
[3] USDA, U.S. Department of Agriculture, Agricultural Research Service. **2011**.
[4] C. G. Fraga, K. D. Croft, D. O. Kennedy, F. A. Tomás-Barberán, *Food Funct.* **2019**, *10*, 514.
[5] M. S. Donia, M. A. Fischbach, *Science* **2015**, *349*, 1254766.
[6] C. Manach, J. Hubert, R. Llorach, A. Scalbert, *Mol. Nutr. Food Res.* **2009**, *53*, 1303.
[7] R. Yin, H.-C. Kuo, R. Hudlikar, D. Sargsyan, S. Li, L. Wang, R. Wu, A.-N. Kong, *Curr.Pharmacol. Rep.* **2019**, *5*, 332.
[8] H. Herrema, J. H. Niess, *Diabetologia* **2020**, *63*, 2533.
[9] A. Agus, K. Clément, H. Sokol, *Gut* **2021**, *70*, 1174.
[10] G. Williamson, C. D. Kay, A. Crozier, *Comprehens. Rev. Food Sci. Food Saf.* **2018**, *17*, 1054.
[11] R. M. de Ferrars, C. Czank, Q. Zhang, N. P. Botting, P. A. Kroon, A. Cassidy, C. D. Kay, *Br. J. Pharmacol.* **2014**, *171*, 3268.
[12] J. Martel, D. M. Ojcius, Y.-F. Ko, J. D. Young, *Trends Biochem. Sci.* **2020**, *45*, 462.
[13] T. Rousu, J. Herttuainen, A. Tolonen, *Rapid Commun. Mass Spectrom.* **2010**, *24*, 939.
[14] J. L. Jiwan, P. Wallemacq, M. F. Hérent, *Clin. Biochem.* **2011**, *44*, 136.
[15] A. F. Ahmad, L. Rich, H. Koch, K. D. Croft, M. G. Ferruzzi, C. D. Kay, J. M. Hodgson, N. C. Ward, *Food Funct.* **2018**, *9*, 6307.
[16] P. J. Curtis, V. van der Velpen, L. Berends, A. Jennings, M. Feelisch, A. M. Umpleby, M. Evans, B. O. Fernandez, M. S. Meiss, M. Minnion, J. Potter, A.-M. Minihane, C. D. Kay, E. B. Rimm, A. Cassidy, *Am. J. Clin. Nutr.* **2019**, *109*, 1535.
[17] C. Czank, A. Cassidy, Q. Zhang, D. J. Morrison, T. Preston, P. A. Kroon, N. P. Botting, C. D. Kay, *Am. J. Clin. Nutr.* **2013**, *97*, 995.
[18] R. M. de Ferrars, C. Czank, Q. Zhang, N. P. Botting, P. A. Kroon, A. Cassidy, C. D. Kay, *Br. J. Pharmacol.* **2014**, *171*, 3268.
[19] S. Hazim, P. J. Curtis, M. Y. Schär, L. M. Ostertag, C. D. Kay, A.-M. Minihane, A. Cassidy, *Am. J. Clin. Nutr.* **2016**, *103*, 694.
[20] D. Nieman, C. Kay, A. Rathore, M. Grace, R. Strauch, E. Stephan, C. Sakaguchi, M. Lila, *Nutrients* **2018**, *10*, 1718.
[21] D. C. Nieman, N. D. Gillitt, G. Y. Chen, Q. Zhang, W. Sha, C. D. Kay, P. Chandra, K. L. Kay, M. A. Lila, *Front Nutr.* **2020**, *7*, 121.

 © 2022 Wiley-VCH GmbH

[22] D. C. Nieman, S. Ramamoorthy, C. D. Kay, C. L. Goodman, C. R. Capps, Z. L. Shue, N. Heyl, M. H. Grace, M. A. Lila, *J. Proteome Res.* **2017**, *16*, 2924.
[23] M. Y. Schär, P. J. Curtis, S. Hazim, L. M. Ostertag, C. D. Kay, J. F. Potter, A. Cassidy, *Am. J. Clin. Nutr.* **2015**, *101*, 931.
[24] S. Bowman, J. Clemens, J. Friday, N. Schroeder, M. Shimizu, R. LaComb, A. Moshfegh Food Patterns Equivalents Intakes by Americans: What We Eat in America, NHANES 2003-2004 and 2015-2016. Food Surveys Research Group. Dietary Data Brief No, 20, 2018.
[25] J. A. Kirlin, FoodAPS National Household Food Acquisition and Purchase Survey. **2013**.
[26] Y. Y. Li, R. Ghanbari, W. Pathmasiri, S. McRitchie, H. Poustchi, A. Shayanrad, G. Roshandel, A. Etemadi, J. D. Pollock, R. Malekzadeh, S. C. J. Sumner, *Front Nutr.* **2020**, *7*, 584585.
[27] Y. Y. Li, C. Douillet, M. Huang, R. Beck, S. J. Sumner, M. Styblo, *Arch. Toxicol.* **2020**, *94*, 1955.
[28] K. H. Liu, M. Nellis, K. Uppal, C. Ma, V. Tran, Y. Liang, D. I. Walker, D. P. Jones, *Anal. Chem.* **2020**, *92*, 8836.
[29] J. Chong, O. Soufan, C. Li, I. Caraus, S. Li, G. Bourque, D. S. Wishart, J. Xia, *Nucleic Acids Res.* **2018**, *46*, W486.
[30] A. Fabregat, S. Jupe, L. Matthews, K. Sidiropoulos, M. Gillespie, P. Garapati, R. Haw, B. Jassal, F. Korninger, B. May, *Nucleic Acids Res.* **2018**, *46*, D649.
[31] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, K. Morishima, *Nucleic Acids Res.* **2017**, *45*, D353.
[32] Y.-Q. Qiu, in *Encyclopedia of Systems Biology* (Eds: W. Dubitzky, O. Wolkenhauer, K.-H. Cho, H. Yokota), Springer New York, New York, NY **2013**, pp. 1068–1069.
[33] D. Zimmer, *Bioanalysis* **2014**, *6*, 13.
[34] N. Semmar, *Metabotype Concept: Flexibility, Usefulness and Meaning in Different Biological Populations*, InTech, **2012**.
[35] P. Lucci, J. Saurina, O. Núñez, *TrAC Trends Anal. Chem.* **2017**, *88*, 1.
[36] L. W. Sumner, A. Amberg, D. Barrett, M. H. Beale, R. Beger, C. A. Daykin, T. W. Fan, O. Fiehn, R. Goodacre, J. L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A. N. Lane, J. C. Lindon, P. Marriott, A. W. Nicholls, M. D. Reily, J. J. Thaden, M. R. Viant, *Metabolomics* **2007**, *3*, 211.
[37] P. C. Hollman, *Arch. Biochem. Biophys.* **2014**, *559*, 100.
[38] D. Del Rio, A. Rodriguez-Mateos, J. P. Spencer, M. Tognolini, G. Borges, A. Crozier, *Antioxid. Redox Signaling* **2013**, *18*, 1818.
[39] P. C. Hollman, A. Cassidy, B. Comte, M. Heinonen, M. Richelle, E. Richling, M. Serafini, A. Scalbert, H. Sies, S. Vidry, *J. Nutr.* **2011**, *141*, 989S.
[40] P. Hajieva, *Molecules* **2017**, *22*, 159.

© 2022 Wiley-VCH GmbH